A posteriori mathematics?

Sharon Berry

July 26, 2009

1 Introduction

Are all mathematical facts a priori? I will argue that, if two common (but esoteric sounding) assumptions hold, then there are (very likely) a posteriori mathematical truths. The two assumptions are: that there are true, but a priori unknowable, pi 01 sentences, and that Malament-Hogarth space times are metaphysically possible. If you accept these claims, there turns out to be a simple route to the conclusion that there are mathematical facts which, not only *can* be learned empirically, but actually *can't* be learned any other way.

2 A posteriori knowledge of mathematics vs. a posteriori mathematical facts

Let me start by considering some more ordinary ways in which our knowledge of mathematics can depend on experience.

Everyone agrees that we can learn mathematical truths 'from experience' in a loose sense: you can learn mathematical truths by hearing testimony, or by looking at a calculator. Arguably, knowledge by testimony has some special features [1], so let me focus on the case of forming new mathematical beliefs by looking at a calculator.

I am going to argue that one's justification for believing mathematical facts learned from a calculator is a posteriori. In particular, if someone learns that, say, the square root of 37 is 6.08276253, by looking at a calculator, their justification for this belief depends on a posteriori beliefs about how the calculator was constructed, the laws of electricity etc. Thus, their belief is not only caused by sensory experience, but depends on these intermediate scientific beliefs for justification. As a result, it too, counts a posteriori.

To simplify things, let's suppose that you designed and built the calculator (so there's no question of the the calculator transmitting testimony from the person who built it to you). You know that a certain algorithm correctly calculates square roots, and you also know (empirically) certain things about electronics: how current would flow through certain wires if certain buttons were pushed, and how this would lead to certain light up displays. Combining these two pieces of knowledge you have built a system where the facts about what the calculator *would* display if certain buttons were pushed systematically covary with the facts about *what the square root function yields* when applied to certain numbers. Thus, you have strong reason to believe that if you push the buttons labeled (say), "sqrt", "37", and then "=", the result will be an inscription of the first 9 digits of the squareroot of 37. Thus, when you *do* push these buttons and the number "6.08276253" shows up, you are justified in believing that the answer will be correct.

Now, note that your scientific beliefs about how the wires in the calculator will behave, play a crucial role in justifying your belief. If your beliefs about the electronics of the calculator were wrong (say the machine is wired up to always print out "6.08276253" when asked for the square root of something, but that

[1] Burge CITE

just happens to be the right answer in this first case you are trying) you would not count as knowing that the square root of 37 is 6.08276253. At best you would be in a Gettier case, believing a true claim about math for reasons that depend on justified, but false, beliefs about how the machine in front of you works. Thus (I claim), your justification that 37 is 6.08276253 depends on your scientific beliefs about the machine, and these in turn, depend on experience.

This is already an interesting - and maybe under-appreciated - fact. (For example, if Burge's motivation in taking testimony to have a distinctive epistemic status (it's supposed to directly transfer justification from one mind to another), is to avoid the conclusion that there can be a posteriori knowledge of mathematical facts, this motivation is undermined by the above-mentioned reasoning about calculators. For, as mentioned above, there's no testimony involved in this case. The person who built the calculator (your earlier self) didn't have the knowledge that 37 is 6.08276253, so they certainly didn't transmit this knowledge to you by way of the calculator. And yet, surely, you would count as having knowledge. So, we will still have to admit that there can be a posteriori knowledge - whatever we say about the status testimony.)

But it falls fall short of the claim that there are a posteriori mathematical *facts*. For, a proposition is a priori iff it can be justified without appeal to experience. And, in the case of the squareroot of 37, someone could (in principle) go through the whole algorithm using pencil and paper, or just reasoning in their head - and thus acquire a justification for this belief about the square root which would not depend on experience.

In contrast, I will argue that our two assumptions (malament-hogarth space times are metaphysically possible, and there are a priori unknowable pi 01 sentences) yield the stronger conclusion, that there are mathematical claims which could be known a posteriori *but could not be known a priori.* That is, not only is there a posteriori knowledge of mathematical facts, but there are a posteriori mathematical facts.

3 Strategy

Now, let's start in on the more ambitious project. My strategy has two steps. The first step is to show how our two assumptions combine to yield the conclusion that it would be metaphysically possible to build a machine which behaves as follows: it turns on a light within 5 minutes if a certain a priori unknowable mathematical statement S is false, and leaves the light unilluminated if the statement was true. The second step (philosophically harder, but crucial step) will be to argue that a creature in this situation could rationally come to believe that they *had* built such a machine, and hence justifiably form the belief that S when they saw the light remain dark after 5 minutes. If the latter claim is true, it follows that S is knowable a posteriori, but not knowable a priori. Thus, S is a genuinely a posteriori mathematical fact.

Note that this is not to say that *we* could learn S a posteriori. We, presumably, do not live in a Malamet-Hogarth space time [2]. So, it might well be that S is permanently unknowable to us. For, S cannot be justified a priori, and the physics of our world (apparently) cannot be shaped to systematically reflect the relevant domain of mathematical facts, in a way that would allow us to learn that S empirically. Nonetheless, S is still a posteriori - if the laws of physics allowed, we could perform an experiment which would let us empirically determine that S, and S cannot be learned without appeal to experience.

[2] Though actually there's some controversy in the physics community over whether similar computers could actually be built CITE

4 The metaphysically possible machine

Our first assumption was that there are some true, but a priori unknowable, pi 01 sentences. Pi 01 sentences are sentences of the form $/forall$ x F(x) where the quantifier ranges over the numbers, and the property F is one that can be recursively checked i.e. you could program a computer to check whether any given number has the property or lacks it, and this computer would be certain to return one answer or the other in a finite amount of time. So, pi 01 sentences say 'every number has the property F' where F is some property which it's easy to check whether any particular number has.

Why think that there are true, but a priori unknowable, pi 01 sentences? It's by no means clear that there are such sentences, but it's commonly assumed that there are, for the following reason. *If* you think that the whole system of justified a priori reasoning can be recursively axiomatized, then the first Incompleteness Theorem yields the conclusion that there are true but a priori unknowable pi 01 sentences. The first Incompleteness theorem says that that no recursively enumerable system, which captures certain basic facts about arithmetic, can prove all true pi 01 sentences. In particular, it shows that that every such system S won't be able to prove the pi 01 sentence 'every number fails to code for a proof of G', where G is a specially-cooked up claim about arithmetic called the Godel sentence for S. Thus *if* there is a definite body of justified a priori reasoning, which could be formalized in such a way that a computer could in principle be programed to check whether any given inference was acceptable, *then* there will be a true pi 01 sentence a priori reasoning cannot lead one to. Thus, there will be a true pi 01 sentence, which cannot be known a priori.

Now, I am going to argue that his claim combines with our second hypothesis to yield the conclusion that it would be metaphysically possible to build a machine that lights up within a specific 5 minute window, if and only if a certain a priori unknowable mathematical statement was true. The first hypothesis says there's at least one a true pi 01 sentence that can't be arrived at via justified a priori reasoning, and hence is unknowable a priori. So let's call this sentence S. S says that every number has a certain property P. And we can check, with pen and paper that, say, 3 has the property P, or with an ordinary computer that the first billion numbers have property P. But S makes a claim about infinitely many different numbers, so no finite amount of checking will suffice. At any finite time, all we know is that we haven't come up with a counter-example yet. Normally, we would try to come up with a mathematical argument that every number has to have property P, for example, by using mathematical induction. But, by hypothesis, there is no justified a priori argument which leads to the conclusion that S.

What we'd like to do would be to, somehow, check infinitely many cases (or build a machine that could do that). And this is where the second hypothesis - that Malement Hogarth space times are metaphysically possible - comes in. The point of these space times, is that they allow a person and an ordinary computer to take different paths through space time, in such a way that information from infinitely many computations in the computer reach the person within what is (for them) a finite amount of time.

Hogarth draws attention to these space times, because a computer user who took such a path relative to the computer would be able to "compute" things that a Turing machine cannot. In particular, a computer user in the situation above would be able to determine whether a regular turing machine halted. They would just set the computer at the other end to go through each stage of the computation, and then send a signal at stage n if and only if the computer halted at stage n. Thus, in particular, they could evaluate any pi 01 sentence (which says that every number has the recursively checkable property F), by programing the regular computer at the other end to go throgh the numbers in order, checking each and sending a signal only if it found a counter example.

Hogarth introduces these kinds of machines because he thinks that computer + user systems of the kind above would count as doing computations, so that whether or not this kind of set up is physically possible will make a difference to what is and isn't computable. Thus, (he argues) it's a contingent matter of physics whether the church turing thesis (that something is computable iff some Turing machine computes it) is false. But for our purposes it doesn't matter whether the opporations in question would count as computations, or whether they are physically possible. Our second hypothesis is simply that it the set -up Hogarth describes is metaphysically possible.

Hoagarth's setup involves two things. A space time, and something that behaves like a Turing machine, that is, like a regular computer minus the fact that a regular computer would eventually break down. Suppose that both of these are metaphysically possible, and can be combined (I take it that the metaphysical possibility of the computer that goes through "infinitely many stages" of computation without breaking is uncontroversial). Then it would be possible to have such a computer, set to check for counter-examples to our unknowable pi 01 sentence S, and send a signal if and only if it found a counter example. If the computer were programed to check instances of some false pi 01 sentence, there would be some stage at which the computer found a counter example, and sent a signal. Let's say the computer sends the signal to some kind of panel which lights up if it ever gets a signal, but otherwise stays unlit. Thus, if the panel stays unlit, whatever pi 01 sentence which the computer was programed to check is true.

Thus we have our first stage: it would be metaphysically possible to build a machine which would turn a light on if an a priori unknowable mathematical statement S is true, and not if this sentence is false.

5 Knowledge of the metaphysically possible machine

But this doesn't yet suffice to show that S is knowable a priori. If a subject knew that he was dealing with a machine of the kind just described, and he saw that the light did not turn on, he would be justified in believing that S sentence was true. But could someone ever know that they were in the kind of set up Hogarth describes?

Establishing this point (that a rational subject could *learn* that the physical system would behave a certain way if S is true) is quite crucial to the argument, and the difficulty of establishing this point explains why I've chosen such a complicated example.

If we just wanted to show that it would be metaphysically possible for there to be a system whose state systematically reflected a priori unknowable mathematical facts, no elaborate description is needed to establish this. It would clearly be metaphysically possible for there to be an oracle which simply spit out the right answer to any mathematical query. There could just be some black box, and a fundamental law of physics which said that that's how the box behaved. But here it's not at all clear what *evidence* could rationally convince you that this really was an oracle. You could check some finite number of predictions from the box, but simply knowing that some physical system gets all the cases you have checked right doesn't justify (if you know nothing more about the system) the conclusion that it gets everything right.

In fact, even more simply, the literal requirement above simply said that the system had to go into a certain state iff a single target sentence S was true. But, actually, *every* object has that property. Since (by stipulation) S is true, ripe bananas have the property of being yellow if and only f S is true, rocks have the property of being attracted to the earth iff S is true etc. But, considering just these cases, there is no reason to think someone could ever learn that the above biconditional holds. (By construction of S, they can't directly learn that S is true by a priori reasoning, and then infer these sentences of the form, if X then S.)

So, even if it's obvious that it would be metaphysically possible to have a system which went into a certain state if and only if a certain pi 01 sentence was true, it's not clear that you could be in a position to know this. And this is what the exotic machinery of experiments in a malament-hogarth space time does for us. For, I will argue, one could get sufficient empirical evidence to convince you that you were in a malamet-hogarth space time, and stood in the relevant relationship to a computer that would check all cases of S?

Admittedly, there is no course of experience which logically entails that the physical world around one is set up in the relevant way. But this is not the standard which we generally accept for knowledge of scientific facts like the structure of space time, or the future behavior of a certain computer. I take the fact that there are some debates about whether the best laws of physics as currently known are compatible with our being in a Malamet-Hogarth space time to be strong evidence that some course of experience could rationally lead one to believe this, and I take the history of physics to provide examples of the kind of evidence which might rationally convince someone that spacetime had a certain structure.

However, even granting that the relevant space time is possible, there might be worries about the turing-machine-realizing computer needed to create Hogarth's set up. Certainly a computer which goes through each stage of the computation and never breaks is metaphysically possible, but would it ever be rational to believe that one was dealing with such a machine? Here I want to suggest that it could look like a simple consequence of basic physical laws (the kind one would rationally expect to be preserved forever) that the turing-computer behaved like this. We might have to imagine evidence that (contrary to what's actually the case) the universe allowed something like perpetual motion, and that energy could be created, but it does seem conceivable that one's best theory of the behavior of some kind of fundamental particle would be that it emits light n second after being hit with a such-and-such wave for the first time, iff the program which checks instances of S halts.

In light of these considerations, it seems like a) someone could be in the situation that Hogarth describes and b) they could come to know that they were in this situation. Furthermore, if they were in this situation, and they saw that the relevant light didn't come on, they could rationally infer from this experience and the fact that they were in the relevant Hogarth set up, that S was true. Thus, it would be possible for someone to learn that S a posteriori. So, if our two hypotheses are true, there can be a priori unknowable facts, which nonetheless could be learned a posteriori (were contingent facts about nature to prove suitably cooperative).

6 Conclusion

In this paper, I have argued that if a) there are a priori unknowable pi 01 sentences and b) Hogarth's set up is metaphysically possible, then there will be a posteriori mathematical facts. That is: there will be mathematical facts which could (in principle) be known, but only in ways that would depend on experience for their justification.